Tien-Tso Ning

14 November, 2019

Linguistics TP5

### 1. Language Modeling

#### 1. Formulate the language model problem for the sentence:

#### 2. Decompose the language model for sentence in (2) using the chain rule.

#### 3. Decompose the language model for the sentence in (2) using the Markov-Assumption

#### 4. Estimate the probability of the sentence in (2) using the Markov decomposition, maximum likelihood estimates and corpus in (1) for training.

25 total words in (2)

Word, Frequency, Maximum Likelihood Estimate

We 2 0.08

identify 1 0.04

remaining 2 0.08

gaps 2 0.08

in 2 0.08

knowledge 3 0.12

. 2 0.08

want 1 0.04

to 1 0.04

boost 1 0.04

their 2 0.08

level 1 0.04

, 1 0.04

get 1 0.04

feedback 1 0.04

on 1 0.04

the 1 0.04

(reminder that the probability of )

#### 5. Estimate the probability of the sentence in (2) using the Markov decomposition, maximum likelihood estimate with Jelinek-Mercer smoothing (assume and corpus (1) for training.